Arboles de Decisión

### Entrega: Ejercicio Base de AD

#### Autor: Ing. Julian Quimbayo Castro

#### Desarrollado con Fecha: 2021-04-14

Paso 1: Importar los Datos

library(readr)  
credit <- read\_csv("C:/Users/PC/Downloads/credit.csv")  
head(credit)

## # A tibble: 6 x 21  
## checking\_balance months\_loan\_duration credit\_history purpose amount  
## <chr> <dbl> <chr> <chr> <dbl>  
## 1 < 0 DM 6 critical radio/tv 1169  
## 2 1 - 200 DM 48 repaid radio/tv 5951  
## 3 unknown 12 critical education 2096  
## 4 < 0 DM 42 repaid furniture 7882  
## 5 < 0 DM 24 delayed car (new) 4870  
## 6 unknown 36 repaid education 9055  
## # ... with 16 more variables: savings\_balance <chr>, employment\_length <chr>,  
## # installment\_rate <dbl>, personal\_status <chr>, other\_debtors <chr>,  
## # residence\_history <dbl>, property <chr>, age <dbl>, installment\_plan <chr>,  
## # housing <chr>, existing\_credits <dbl>, default <dbl>, dependents <dbl>,  
## # telephone <chr>, foreign\_worker <chr>, job <chr>

Paso 2: Revisar las estructuras de los datos

str(credit)

## spec\_tbl\_df[,21] [1,000 x 21] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ checking\_balance : chr [1:1000] "< 0 DM" "1 - 200 DM" "unknown" "< 0 DM" ...  
## $ months\_loan\_duration: num [1:1000] 6 48 12 42 24 36 24 36 12 30 ...  
## $ credit\_history : chr [1:1000] "critical" "repaid" "critical" "repaid" ...  
## $ purpose : chr [1:1000] "radio/tv" "radio/tv" "education" "furniture" ...  
## $ amount : num [1:1000] 1169 5951 2096 7882 4870 ...  
## $ savings\_balance : chr [1:1000] "unknown" "< 100 DM" "< 100 DM" "< 100 DM" ...  
## $ employment\_length : chr [1:1000] "> 7 yrs" "1 - 4 yrs" "4 - 7 yrs" "4 - 7 yrs" ...  
## $ installment\_rate : num [1:1000] 4 2 2 2 3 2 3 2 2 4 ...  
## $ personal\_status : chr [1:1000] "single male" "female" "single male" "single male" ...  
## $ other\_debtors : chr [1:1000] "none" "none" "none" "guarantor" ...  
## $ residence\_history : num [1:1000] 4 2 3 4 4 4 4 2 4 2 ...  
## $ property : chr [1:1000] "real estate" "real estate" "real estate" "building society savings" ...  
## $ age : num [1:1000] 67 22 49 45 53 35 53 35 61 28 ...  
## $ installment\_plan : chr [1:1000] "none" "none" "none" "none" ...  
## $ housing : chr [1:1000] "own" "own" "own" "for free" ...  
## $ existing\_credits : num [1:1000] 2 1 1 1 2 1 1 1 1 2 ...  
## $ default : num [1:1000] 1 2 1 1 2 1 1 1 1 2 ...  
## $ dependents : num [1:1000] 1 1 2 2 2 2 1 1 1 1 ...  
## $ telephone : chr [1:1000] "yes" "none" "none" "none" ...  
## $ foreign\_worker : chr [1:1000] "yes" "yes" "yes" "yes" ...  
## $ job : chr [1:1000] "skilled employee" "skilled employee" "unskilled resident" "skilled employee" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. checking\_balance = col\_character(),  
## .. months\_loan\_duration = col\_double(),  
## .. credit\_history = col\_character(),  
## .. purpose = col\_character(),  
## .. amount = col\_double(),  
## .. savings\_balance = col\_character(),  
## .. employment\_length = col\_character(),  
## .. installment\_rate = col\_double(),  
## .. personal\_status = col\_character(),  
## .. other\_debtors = col\_character(),  
## .. residence\_history = col\_double(),  
## .. property = col\_character(),  
## .. age = col\_double(),  
## .. installment\_plan = col\_character(),  
## .. housing = col\_character(),  
## .. existing\_credits = col\_double(),  
## .. default = col\_double(),  
## .. dependents = col\_double(),  
## .. telephone = col\_character(),  
## .. foreign\_worker = col\_character(),  
## .. job = col\_character()  
## .. )

## Relación entre cuenta de ahorros y corriente  
table(credit$checking\_balance)

##   
## < 0 DM > 200 DM 1 - 200 DM unknown   
## 274 63 269 394

table(credit$savings\_balance)

##   
## < 100 DM > 1000 DM 101 - 500 DM 501 - 1000 DM unknown   
## 603 48 103 63 183

## Summary duración del credito vs cantidad de prestamo  
summary(credit$months\_loan\_duration)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4.0 12.0 18.0 20.9 24.0 72.0

summary(credit$amount)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 250 1366 2320 3271 3972 18424

## Variable a predecir  
table(credit$default)

##   
## 1 2   
## 700 300

Paso 3: dataframe y factor 1=si 2=no

class(credit)

## [1] "spec\_tbl\_df" "tbl\_df" "tbl" "data.frame"

credit$default <- ifelse(credit$default ==1, "Si", "No")  
  
table(credit$default)

##   
## No Si   
## 300 700

str(credit)

## spec\_tbl\_df[,21] [1,000 x 21] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ checking\_balance : chr [1:1000] "< 0 DM" "1 - 200 DM" "unknown" "< 0 DM" ...  
## $ months\_loan\_duration: num [1:1000] 6 48 12 42 24 36 24 36 12 30 ...  
## $ credit\_history : chr [1:1000] "critical" "repaid" "critical" "repaid" ...  
## $ purpose : chr [1:1000] "radio/tv" "radio/tv" "education" "furniture" ...  
## $ amount : num [1:1000] 1169 5951 2096 7882 4870 ...  
## $ savings\_balance : chr [1:1000] "unknown" "< 100 DM" "< 100 DM" "< 100 DM" ...  
## $ employment\_length : chr [1:1000] "> 7 yrs" "1 - 4 yrs" "4 - 7 yrs" "4 - 7 yrs" ...  
## $ installment\_rate : num [1:1000] 4 2 2 2 3 2 3 2 2 4 ...  
## $ personal\_status : chr [1:1000] "single male" "female" "single male" "single male" ...  
## $ other\_debtors : chr [1:1000] "none" "none" "none" "guarantor" ...  
## $ residence\_history : num [1:1000] 4 2 3 4 4 4 4 2 4 2 ...  
## $ property : chr [1:1000] "real estate" "real estate" "real estate" "building society savings" ...  
## $ age : num [1:1000] 67 22 49 45 53 35 53 35 61 28 ...  
## $ installment\_plan : chr [1:1000] "none" "none" "none" "none" ...  
## $ housing : chr [1:1000] "own" "own" "own" "for free" ...  
## $ existing\_credits : num [1:1000] 2 1 1 1 2 1 1 1 1 2 ...  
## $ default : chr [1:1000] "Si" "No" "Si" "Si" ...  
## $ dependents : num [1:1000] 1 1 2 2 2 2 1 1 1 1 ...  
## $ telephone : chr [1:1000] "yes" "none" "none" "none" ...  
## $ foreign\_worker : chr [1:1000] "yes" "yes" "yes" "yes" ...  
## $ job : chr [1:1000] "skilled employee" "skilled employee" "unskilled resident" "skilled employee" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. checking\_balance = col\_character(),  
## .. months\_loan\_duration = col\_double(),  
## .. credit\_history = col\_character(),  
## .. purpose = col\_character(),  
## .. amount = col\_double(),  
## .. savings\_balance = col\_character(),  
## .. employment\_length = col\_character(),  
## .. installment\_rate = col\_double(),  
## .. personal\_status = col\_character(),  
## .. other\_debtors = col\_character(),  
## .. residence\_history = col\_double(),  
## .. property = col\_character(),  
## .. age = col\_double(),  
## .. installment\_plan = col\_character(),  
## .. housing = col\_character(),  
## .. existing\_credits = col\_double(),  
## .. default = col\_double(),  
## .. dependents = col\_double(),  
## .. telephone = col\_character(),  
## .. foreign\_worker = col\_character(),  
## .. job = col\_character()  
## .. )

##Conversión a factor  
credit$default <- factor(credit$default)

Paso 4: Preparar los Datos

set.seed(123456)  
credit\_rand <- credit[order(runif(1000)),]  
  
###Comprobar  
summary(credit$amount)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 250 1366 2320 3271 3972 18424

summary(credit\_rand$amount)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 250 1366 2320 3271 3972 18424

class(credit\_rand)

## [1] "tbl\_df" "tbl" "data.frame"

credit\_rand <- as.data.frame(credit\_rand)

Paso 5: Datos Testing y Train

credit\_train <- credit\_rand[1:900,]  
credit\_test <- credit\_rand[901:1000,]  
  
### Comprobar el 90% train y el 10% test  
prop.table(table(credit\_train$default))

##   
## No Si   
## 0.3 0.7

prop.table(table(credit\_test$default))

##   
## No Si   
## 0.3 0.7

Paso 6: Creación del modelo AD C5.0

##install.packages("C50")  
library(C50)  
credit\_modelo <- C5.0(credit\_train[-17], credit\_train$default)  
  
credit\_modelo

##   
## Call:  
## C5.0.default(x = credit\_train[-17], y = credit\_train$default)  
##   
## Classification Tree  
## Number of samples: 900   
## Number of predictors: 20   
##   
## Tree size: 54   
##   
## Non-standard options: attempt to group attributes

summary(credit\_modelo)

##   
## Call:  
## C5.0.default(x = credit\_train[-17], y = credit\_train$default)  
##   
##   
## C5.0 [Release 2.07 GPL Edition] Wed Apr 14 21:59:26 2021  
## -------------------------------  
##   
## Class specified by attribute `outcome'  
##   
## Read 900 cases (21 attributes) from undefined.data  
##   
## Decision tree:  
##   
## checking\_balance in {unknown,> 200 DM}:  
## :...installment\_plan = none: Si (349/33)  
## : installment\_plan in {stores,bank}:  
## : :...age > 44: Si (16)  
## : age <= 44:  
## : :...purpose in {radio/tv,furniture,retraining,car (used),repairs,  
## : : domestic appliances,others}: Si (31/5)  
## : purpose in {business,car (new),education}:  
## : :...residence\_history <= 1: Si (3)  
## : residence\_history > 1:  
## : :...savings\_balance in {< 100 DM,501 - 1000 DM,unknown,  
## : : 101 - 500 DM}: No (15/1)  
## : savings\_balance = > 1000 DM: Si (2)  
## checking\_balance in {1 - 200 DM,< 0 DM}:  
## :...other\_debtors = guarantor:  
## :...purpose in {business,car (new)}: No (5/1)  
## : purpose in {radio/tv,furniture,education,retraining,car (used),repairs,  
## : domestic appliances,others}: Si (30/2)  
## other\_debtors in {none,co-applicant}:  
## :...months\_loan\_duration > 30:  
## :...employment\_length = unemployed: Si (6)  
## : employment\_length in {4 - 7 yrs,1 - 4 yrs,> 7 yrs,0 - 1 yrs}:  
## : :...savings\_balance in {< 100 DM,> 1000 DM,501 - 1000 DM,  
## : : 101 - 500 DM}: No (78/18)  
## : savings\_balance = unknown:  
## : :...checking\_balance = 1 - 200 DM: Si (9/1)  
## : checking\_balance = < 0 DM: No (3)  
## months\_loan\_duration <= 30:  
## :...credit\_history = critical: Si (86/20)  
## credit\_history = fully repaid this bank:  
## :...property in {building society savings,  
## : : unknown/none}: No (13)  
## : property in {real estate,other}: Si (9/3)  
## credit\_history = delayed:  
## :...property in {building society savings,real estate,  
## : : other}: Si (26/6)  
## : property = unknown/none: No (3)  
## credit\_history = fully repaid:  
## :...housing in {rent,for free}: No (8)  
## : housing = own:  
## : :...age <= 34: Si (8/2)  
## : age > 34: No (3)  
## credit\_history = repaid:  
## :...savings\_balance in {> 1000 DM,501 - 1000 DM}: Si (16/1)  
## savings\_balance in {< 100 DM,unknown,101 - 500 DM}:  
## :...purpose in {retraining,domestic appliances,  
## : others}: No (7/2)  
## purpose = repairs: Si (4/1)  
## purpose = business:  
## :...age <= 35: Si (8)  
## : age > 35: No (4/1)  
## purpose = education:  
## :...checking\_balance = 1 - 200 DM: Si (2)  
## : checking\_balance = < 0 DM: No (4)  
## purpose = car (used):  
## :...amount <= 6967: Si (9/1)  
## : amount > 6967: No (4)  
## purpose = radio/tv:  
## :...employment\_length in {4 - 7 yrs,  
## : : unemployed}: No (4/1)  
## : employment\_length = > 7 yrs: Si (12/2)  
## : employment\_length = 1 - 4 yrs:  
## : :...existing\_credits <= 1: Si (13/4)  
## : : existing\_credits > 1: No (2)  
## : employment\_length = 0 - 1 yrs:  
## : :...foreign\_worker = yes: No (13/4)  
## : foreign\_worker = no: Si (2)  
## purpose = car (new):  
## :...personal\_status in {divorced male,  
## : : married male}: No (6/1)  
## : personal\_status = female:  
## : :...months\_loan\_duration <= 10: Si (2)  
## : : months\_loan\_duration > 10: No (17/1)  
## : personal\_status = single male:  
## : :...savings\_balance in {unknown,  
## : : 101 - 500 DM}: Si (8/1)  
## : savings\_balance = < 100 DM:  
## : :...checking\_balance = 1 - 200 DM: Si (6/1)  
## : checking\_balance = < 0 DM: No (9/2)  
## purpose = furniture:  
## :...job = unskilled resident: Si (8/2)  
## job = unemployed non-resident: No (1)  
## job = mangement self-employed:  
## :...amount <= 3181: Si (2)  
## : amount > 3181: No (2)  
## job = skilled employee:  
## :...employment\_length = 0 - 1 yrs: Si (5/1)  
## employment\_length = unemployed: No (2)  
## employment\_length = 4 - 7 yrs:  
## :...months\_loan\_duration <= 22: Si (3)  
## : months\_loan\_duration > 22: No (2)  
## employment\_length = > 7 yrs:  
## :...savings\_balance in {< 100 DM,  
## : : 101 - 500 DM}: Si (3)  
## : savings\_balance = unknown: No (2)  
## employment\_length = 1 - 4 yrs:  
## :...checking\_balance = 1 - 200 DM: No (4)  
## checking\_balance = < 0 DM:  
## :...age <= 23: No (4)  
## age > 23: Si (7/1)  
##   
##   
## Evaluation on training data (900 cases):  
##   
## Decision Tree   
## ----------------   
## Size Errors   
##   
## 54 119(13.2%) <<  
##   
##   
## (a) (b) <-classified as  
## ---- ----  
## 183 87 (a): class No  
## 32 598 (b): class Si  
##   
##   
## Attribute usage:  
##   
## 100.00% checking\_balance  
## 53.78% other\_debtors  
## 49.89% months\_loan\_duration  
## 46.22% installment\_plan  
## 39.22% credit\_history  
## 33.78% savings\_balance  
## 29.67% purpose  
## 19.33% employment\_length  
## 11.22% age  
## 5.67% property  
## 5.33% personal\_status  
## 5.00% job  
## 2.22% residence\_history  
## 2.11% housing  
## 1.89% amount  
## 1.67% existing\_credits  
## 1.67% foreign\_worker  
##   
##   
## Time: 0.0 secs

Paso 7: Matriz de Confusión

credit\_pred <- predict(credit\_modelo, credit\_test)  
  
library(gmodels)  
CrossTable(credit\_test$default, credit\_pred, prop.chisq = FALSE, prop.c = FALSE,   
 prop.r = FALSE, dnn = c('Default Actual', 'Default Predecido'))

##   
##   
## Cell Contents  
## |-------------------------|  
## | N |  
## | N / Table Total |  
## |-------------------------|  
##   
##   
## Total Observations in Table: 100   
##   
##   
## | Default Predecido   
## Default Actual | No | Si | Row Total |   
## ---------------|-----------|-----------|-----------|  
## No | 11 | 19 | 30 |   
## | 0.110 | 0.190 | |   
## ---------------|-----------|-----------|-----------|  
## Si | 15 | 55 | 70 |   
## | 0.150 | 0.550 | |   
## ---------------|-----------|-----------|-----------|  
## Column Total | 26 | 74 | 100 |   
## ---------------|-----------|-----------|-----------|  
##   
##

Tablas

library(DT)  
datatable(head(iris,80),class = 'cell-border stripe',options = list(  
 pageLength = 5, autoWidth = TRUE  
))